

Coupled virus - bacteria interactions and ecosystem function in an engineered microbial system

M.R. Brown ^{a,*}, J.C. Baptista ^{a,1}, M. Lunn ^b, D.L. Swan ^{a,2}, S.J. Smith ^{a,3}, R.J. Davenport ^a, B.D. Allen ^a, W.T. Sloan ^c, T.P. Curtis ^a

^a School of Engineering, Newcastle University, NE1 7RU, UK

^b Department of Statistics, University of Oxford, OX1 3TG, UK

^c Department of Civil Engineering, University of Glasgow, G12 8LT, UK

ARTICLE INFO

Article history:

Received 18 September 2018

Received in revised form

6 January 2019

Accepted 7 January 2019

Available online 11 January 2019

Keywords:

Viruses

Activated sludge

Wastewater treatment

ABSTRACT

Viruses are thought to control bacterial abundance, affect community composition and influence ecosystem function in natural environments. Yet their dynamics have seldom been studied in engineered systems, or indeed in any system, for long periods of time. We measured virus abundance in a full-scale activated sludge plant every week for two years. Total bacteria and ammonia oxidising bacteria (AOB) abundances, bacterial community profiles, and a suite of environmental and operational parameters were also monitored. Mixed liquor virus abundance fluctuated over an order of magnitude (3.18×10^8 – 3.41×10^9 virus's mL⁻¹) and that variation was statistically significantly associated with total bacterial and AOB abundance, community composition, and effluent concentrations of COD and NH₄⁺-N and thus system function. This suggests viruses play a far more important role in the dynamics of activated sludge systems than previously realised and could be one of the key factors controlling bacterial abundance, community structure and functional stability and may cause reactors to fail. These findings are based on statistical associations, not mechanistic models. Nevertheless, viral associations with abiotic factors, such as pH, make physical sense, giving credence to these findings and highlighting the role that physical factors play in virus ecology. Further work is needed to identify and quantify specific bacteriophage and their hosts to enable us to develop mechanistic models of the ecology of viruses in wastewater treatment systems. However, since we have shown that viruses can be related to effluent quality and virus quantification is simple and cheap, practitioners would probably benefit from quantifying viruses now.

© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Viruses are agents of mortality (Wommack and Colwell, 2000), nutrient regeneration (Middelboe and Jorgensen, 2006; Haaber and Middelboe, 2009; Shelford et al., 2012) and horizontal gene transfer (Lindell et al., 2004; Sullivan et al., 2006) and therefore key drivers of bacterial abundance, activity and community composition in natural environments, as well ecosystem function (Rodriguez-

Valera et al., 2009; Winter et al., 2010; Breitbart, 2012; Liu et al., 2015a). Yet their dynamics have seldom been monitored in activated sludge or indeed any other engineered microbial ecosystem.

Recent advances in molecular methods and the adoption of ecological approaches to engineered systems have shed some light on the complex mechanisms driving the microbial communities in, and the performance and functional stability of, activated sludge systems. Deterministic (such as reactor design, process configuration, operational and environmental conditions) and stochastic processes (microbial birth, death, and immigration), are thought to be at work (van der Gast et al., 2008; Wells et al., 2009, 2011; Ayarza et al., 2010; Ofiteru et al., 2010; Ayarza and Erijman, 2011; Valentín-Vargas et al., 2012). Yet, neither completely explain the variation seen in the microbial composition and performance of such systems (Ofiteru et al., 2010).

Given the prominent role of viruses in the microbial ecology of natural environments viral infection could be another important

* Corresponding author.

E-mail address: mathew.brown@ncl.ac.uk (M.R. Brown).

¹ Department for Work and Pensions, Benton Park View, Newcastle upon Tyne, NE1 98YX, UK.

² School of Engineering, Merz Court, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.

³ Wellcome Trust Centre for Mitochondrial Research, Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

mechanism, especially considering its connection with host abundance fluctuations and functional instability (Lee et al., 2007; Barr et al., 2010; Motlagh et al., 2015), evidence of predator-prey type oscillations (Lee et al., 2007; Otawa et al., 2007) and the sheer abundance of viruses (Otawa et al., 2007; Wu and Liu, 2009; Brown et al., 2015) within activated sludge systems. To date however, viral dynamics have been largely overlooked.

The absence of methods to link viruses to their hosts and couple their abundances (Dang and Sullivan, 2014; Brum and Sullivan, 2015) is undoubtedly a contributing factor. However, even total abundance methods, which sparked a transformation in virus ecology in the 1980s (Bergh et al., 1989) and 1990s (Hara et al., 1991; Marie et al., 1999), have only found limited application in activated sludge systems: being used to compare different activated sludge plants (Wu and Liu, 2009; Brown et al., 2015), or across very modest time scales (Otawa et al., 2007). This is in contrast to the numerous studies that have yielded great insights into, and underpin our understanding of, virus ecology in marine (e.g. Jiang and Paul (1994), Weinbauer et al. (1995), Bratbak et al. (1996) and Li and Dickie (2001)) and freshwater environments (e.g. Hennes and Simon (1995), Hofer and Sommaruga (2001), Bettarel et al. (2004) and Jacquet et al. (2005)). Although even here the spatial and temporal scale of such studies are typically modest; a consequence of limited “ship time” and other logistical constraints. Thus multi-year studies at functionally relevant temporal scales, and which incorporate time varying exogenous factors, have recently been called for in viral ecology (Breitbart, 2012; Brum and Sullivan, 2015; Wigington et al., 2016).

To this end we used a recently adapted flow cytometry (FCM) method (Brown et al., 2015) to measure virus abundance weekly for two years in a full scale nitrifying activated sludge plant. The sampling frequency and duration was considered logistically feasible and biologically and functionally relevant considering nitrifying activated sludge systems maintain solid retention times (SRT) > 7 days (the average period of time biomass remains within a system, Tchobanoglous et al. (2003)). The relationship between virus abundance and the dynamics of total and ammonia oxidising bacteria (AOB), as well as community structure, was evaluated, and the influence of exogenous factors (environmental and operational parameters) was also examined. The role of viruses in the microbial ecology, and performance (removal of COD and $\text{NH}_4^+ - \text{N}$) of activated sludge systems was established.

2. Materials and methods

2.1. Sample collection

Mixed liquor (ML) grab samples were collected from the aeration basin (3600 m^3) of a conventional nitrifying domestic wastewater treatment plant (WWTP, $6751 \text{ m}^3 \text{ day}^{-1}$), situated in the North East of England, United Kingdom, on a weekly basis for a period of two years from June 2011 to May 2013 (104 weeks). Samples were collected in 50 mL polypropylene containers and transported to the lab on ice for immediate processing. Concurrent primary settled sewage (influent) and effluent samples were also collected, in addition to a number of operational variables.

2.2. Analytical methods

For all samples (influent, ML and effluent) suspended/volatile suspended solids (SS/VSS), soluble chemical oxygen demand (COD_s) and soluble ammonium ($\text{NH}_4^+ - \text{N}$) were determined according to Standard Methods (APHA, 1998) and using Merck COD and $\text{NH}_4^+ - \text{N}$ test kits (VWR, UK) respectively. Anion concentrations, including nitrate, nitrite, sulphate and phosphate, were determined

using high performance Ion Chromatography (Dionex ICS-1000 with AS40 auto sampler); samples were filtered through a $0.2 \mu\text{m}$ polyethersulfone membrane prior to analysis. Influent trace metals, including cadmium, zinc, lead and copper, were measured by inductively coupled plasma optical emission spectroscopy (ICP-OES) (Vista MPX axial ICP-OES, Varian, UK), as described by Martin et al. (1994). Samples were acidified on collection to $\text{pH} < 2$, digested and then filtered through a $0.45 \mu\text{m}$ polyethersulfone membrane prior to analysis. Finally temperature, dissolved oxygen (DO) and pH within the aeration basin were measured in real time using *in situ* probes, influent flow rate was determined using a Parshall flume, and sludge age was obtained from plant operators.

2.3. Molecular methods

2.3.1. Flow cytometry

For virus enumeration 1 mL sub-samples of influent, ML and effluent were taken, transferred into 2 mL cryovials and fixed at a final concentration of 0.5% glutaraldehyde for 15–30 min at 4°C in the dark. Samples were then flash frozen in liquid nitrogen and stored at -80°C . After defrosting, samples were pre-treated and analysed in triplicate as described by Brown et al. (2015) using a FACScan flow cytometer (Becton Dickinson, USA) equipped with a 15-mW 488-nm air-cooled argon-ion laser and a standard filter setup.

Using this FCM method not all events detected are necessarily viruses, membrane-derived vesicles and gene transfer agents (non-virus particles) could simultaneously be detected and quantified (Forterre et al., 2013; Biller et al., 2017). Hence the term virus like particle (VLP) being coined and used within marine viral ecology. However, in an earlier study (Brown et al., 2015) regression analysis of FCM and transmission electron microscopy virus counts yielded a highly significant, positive correlation coefficient ($P < 0.05$), a high coefficient of determination ($R^2 = 0.77$) and an intercept and slope coefficient statistically indistinguishable from 0 and 1 (not explicitly stated in Brown et al. (2015)) respectively, thus robust statistical evidence that non-virus particles affect FCM virus counts in activated sludge is lacking. Indeed, Biller et al. (2017) concluded that non-virus particles are unlikely to systematically inflate measurements to a notable degree. It is however conceivable, given the R^2 of 0.77 in Brown et al. (2015), that non-virus like particles contribute to random measurement error. Thus in the absence of a systematic measurement error we have elected to use the simpler term virus over VLP.

2.3.2. DNA extraction

DNA was extracted from 250 μL of ML and from 15 mL of influent, the latter being centrifuged at $3392 \times g$ for 15 min and the supernatant removed down to a working volume of 250 μL . Cell wall disruption was then carried out using the FastDNA SPIN Kit for soil (MP Biomedicals, USA), thus 244.5 μL of sodium phosphate buffer and 30.5 μL of MT buffer was added to samples and the mixture transferred to Lysing Matrix E tubes. Samples were then lysed at 6.5 ms^{-1} for 30 s in a FastPrep instrument (MP Biomedicals, USA) and centrifuged at $14000 \times g$ for 15 min. DNA from 250 μL of the supernatant was then purified using a MagNA Pure LC 2.0 (Roche, UK) and the MagNA Pure LC DNA Isolation Kit S.

2.3.3. Illumina sequencing

Sample preparation for Illumina sequencing generally followed the protocol of Caporaso et al. (2012). Thus the V4 region of the bacterial 16S rRNA gene was amplified using primers 515F (aat gatacggcaccacggagatctacactatgttggtgtGTGCCAGCMGCCGCGTAA; adapter, primer pad and primer linker in small letters; and specific primer sequence in capital letters) and 806R (caagcagaagacg gcatacagagatbarcodeagtcagtcagccGGACTACHVGGGTWTCTAAT), the

latter was barcoded with a 12-base error-correcting Golay code to facilitate sample multiplexing. Each sample was amplified in duplicate, pooled and then cleaned using a MinElute 96 UF Purification Kit as per the manufacturer's instructions (Qiagen Ltd., West Sussex, UK). PCR reactions consisted of: 2 µl DNA extract, 20 µl 5 Prime Hot Master Mix (VWR, Lutterworth, UK), 1 µl each of forward and reverse primer (10 µM final concentration), and 26 µl molecular-grade water. Reactions were denatured at 94 °C for 3 min, with amplification proceeding for 35 cycles at 94 °C for 45 s (further denaturing), 50 °C for 60 s (annealing) and 72 °C for 90 s (extension); a final extension at 72 °C for 10 min was added to ensure complete amplification. The concentration and purity of pooled amplicons was assessed using a Quant-iT PicoGreen dsDNA assay kit (Life Technologies, Paisley, UK). A composite sample for sequencing was created by combining all samples in equimolar amounts. The composite sample was cleaned twice using Agen-court AMPure XP beads (Beckman Coulter, High Wycombe, UK) according to the manufacturer's instructions, with fragment selection undertaken using E-Gel 2% agarose gels (Life Technologies, Paisley, UK). The size-selected fragment was then purified using the QIAquick PCR purification kit (Qiagen Ltd., West Sussex, UK) and quantified using the Quant-iT PicoGreen dsDNA assay kit, to ensure the sample contained enough DNA for sequencing analysis. Sequencing was carried out on the Illumina MiSeq personal sequencer at the Centre for Genomic Research, University of Liverpool. A total of 9.5 million reads were obtained.

Raw reads were processed using the DADA2 (v. 1.4, Callahan et al. (2016)) pipeline, specifically following the workflow for Big Data (<https://benjjneb.github.io/dada2/bigdata.html>), and using R version 3.4.0, R Core Team (2017)). Forward and reverse read pairs were trimmed and filtered (minimum length 200 nucleotides, Emax < 2 expected errors). Amplicon sequence variants (ASVs) were independently inferred in each sample from forward and reverse reads using the run-specific error rates, and then joined using the `~mergePairs ~` function. Chimeric ASVs were inferred and identified using `~removeBimeraDenovo~` and removed. This resulted in 2756 final ASVs (per sample average = 20750, min = 8708 and max = 63995) which were taxonomically classified against the SILVA database (v.128, Quast et al. (2013)) using the DADA2 implementation of the RDP's naive Bayesian classifier (Wang et al., 2007). Note ASVs are denoted as operational taxonomic units (OTUs).

2.3.3.1. qPCR. Quantification of total bacteria and AOB was carried out using qPCR and amplification of the 16S rRNA gene and the ammonia monooxygenase (*amoA*) gene respectively. Samples were amplified in triplicate on a CFX96 Real-Time PCR Detection System (Bio-Rad, UK) using the primer sets 338F (Muyzer et al., 1993) and 1046R (Huber et al., 2007) for total bacteria and *amoA*-1F* (Stephen et al., 1999) and *amoA*-2R (Rotthauwe et al., 1997) for AOB. qPCR reactions contained 3 µL of template DNA (sample DNA, standard DNA or molecular grade water (negative control)), 0.5 µL of forward and reverse primer (10 pmoles per µL), 5 µL of SsoFast EvaGreen supermix (Bio-Rad, UK) and 1 µL of molecular-grade water. Reaction conditions were: 1 cycle at 98 °C for 3 min, followed by 40 cycles consisting of 98 °C for 5 s and 60 °C (338F/1046R) or 56 °C (*amoA*-1F/*amoA*-2R) for 5 s. Purified circular plasmids containing the target gene were used as standards and run in triplicate for each qPCR reaction. Efficiencies for all qPCR reactions ranged between 90 and 110% and had a $R^2 \geq 0.99$. Gene copy numbers per unit volume were converted to cell numbers per unit volume using accompanying sequence data for the 16S rRNA gene (described in SI) and assuming each AOB cell contained 2 copies of the *amoA* gene (McTavish et al., 1993; Norton et al., 2002).

2.4. Statistical analysis

All statistical analysis, unless otherwise stated, was undertaken in RStudio (v. 1.0.143, R Core Team (2017)) using R version 3.4.0 (R Core Team, 2017).

2.4.1. Virus – biotic/abiotic interactions

Virus – biotic/abiotic interactions were assessed by multivariate generalised least squares (GLS) regression (`gls`, “nlme” v. 3.1–131, Pinheiro et al. (2017)), with models describing mixed liquor virus, total bacteria and AOB abundance produced. All measured biotic and abiotic parameters, unless otherwise stated (see footnotes of Table S2 – S4), were initially used as covariates ($K_i = \sim 59$, $n = 102$), with multivariate ordinary least squares (OLS) regression (`lm`, “stats” v. 3.4.0, R Core Team (2017)) followed by bidirectional elimination (stepAIC, “MASS” v. 7.3–47, Venables et al. (2002)) based on Bayesian information criterion (BIC) used for model simplification and selection. OLS models were then rerun as GLS to allow incorporation of a correlation structure (`corCAR1(form = ~ Week)`, accounting for slight variations in sampling frequency (approximately weekly). To guarantee parsimony and adherence to the assumptions of linear regression manual backward elimination and forward selection based on BIC was subsequently undertaken on all GLS models, which were fit by maximum likelihood.

Models were checked visually for linearity (Fig. S4 – S6, A and B), homoscedasticity (Fig. S4 – S6, B and C), residual autocorrelation (`acf` and `pacf`, “stats” v. v. 3.4.0, R Core Team (2017); Fig. S4 – S6, E and F), and normality (`qqPlot`, “car” v. 2.1–4, Fox and Weisberg (2011); Fig. S4 – S6, D), with the Anderson-Darling Test ($P > 0.05$, `ad.test`, “nortest” v. 1.0–4, Gross and Ligges (2015)) additionally used to confirm the latter. Collinearity amongst explanatory variables was assessed by variance inflation factor's (VIF's, `vif`, “car” v. 2.1–4, Fox and Weisberg (2011)), with variables contributing to VIF's > 3 removed based on statistical significance (P -value) until all fell below this threshold (Zuur et al., 2010). To aid in adherence to these assumptions variables were either transformed (\log_{10} for virus, total bacteria and AOB abundances (biotic) and natural log for COD, NH_4^+ -N, nitrate, nitrite and phosphate terms (abiotic)) or standardised to mean 0 variance 1 (all other environmental and operational parameters (abiotic)). Finally ANOVA (`anova`, “stats” v. v. 3.4.0, R Core Team (2017)) was used to test the statistical significance of each model's correlation structure, whilst a pseudo- R^2 was calculated to ascertain a representation of the variance in mixed liquor abundances explained by each of the models respectively (`r.squaredLR`, “MuMin” v. 1.15.6, Barton (2016)).

2.4.2. Virus – community structure interactions

Canonical correspondence analysis (CCA) was performed to assess the response of the bacterial communities to abiotic/biotic conditions (`cca`, “vegan” v. 2.4–3, Oksanen et al. (2018)), with all measured biotic and abiotic parameters, transformed as in 2.4.1, initially used as explanatory variables ($K_i = \sim 59$, $n = 102$) unless otherwise stated (see caption of Fig. 3). Automated bidirectional selection, based on Akaike information criterion (AIC) and Monte Carlo permutation tests (999 permutations), was then used for model simplification and thus identification of the most significant explanatory variables (step, test = “perm”, “stats” v. v. 3.4.0, R Core Team (2017)). Collinearity was assessed by VIF's (`vif.cca`, “vegan” v. 2.4–3, Oksanen et al. (2018)), with variables contributing to VIF's > 3 being removed as in 2.4.1 (Zuur et al., 2010). Finally the statistical significance of the CCA model (constrained components), its axes and the marginal effects of each explanatory variable were assessed using permutation tests (999 permutations, `anova.cca`, “vegan” v. 2.4–3, Oksanen et al. (2018)). Note CCA was chosen over

redundancy analysis as unimodal approaches are better suited to relative abundances and the presence of zeros (Ramette, 2007).

Local similarity analysis (LSA, Ruan et al. (2006)) was utilised to observe correlations between the 50 most abundant OTUs (relative abundances) and mixed liquor virus abundance. Analysis was undertaken using eLSA v 1.0.2 (Xia et al., 2011, 2013) in Python v 2.7. A maximum time delay of three was utilised (delayLimit = 3), *P*-values were calculated by permutation tests (1000, *P*-valueMethod = perm), the required precision of *P*-values was set at 1/1000 (precision = 1000) and data was rank-normalised and z-transformed (normMethod = robustZ, Ruan et al., 2006; Xia et al., 2011, 2013). Multiple hypothesis correction was undertaken using *Q*-values (Storey and Tibshirani, 2003). The LSA output was then visualised as an association network in Cytoscape v 3.6.0 (Shannon et al., 2003), only correlations with a *P*-value ≤ 0.05 , a LSA score (LS) ≥ 0.3 and a *Q*-value ≤ 0.01 were examined. Calculation of the peak (max relative abundance) to average (mean relative abundance) ratio (PAR) for examined OTUs allowed assessment of their persistence within the mixed liquor, three arbitrarily defined ecological categories facilitated this process; persistent (PAR ≤ 5), intermittent (PAR $> 5 < 10$) and transient (PAR ≥ 10).

The influence of virus abundance on alpha diversity was assessed by calculation of Spearman's rank correlation coefficients (cor.test, "stats" v. 3.4.0, R Core Team, 2017) between mixed liquor and effluent virus abundance and D_1 (exponential of Shannon diversity) and D_2 (inverse of Simpson diversity) Hills diversities (diversity, "vegan" v. 2.4–3, Oksanen et al., 2018), which better represent rare and common taxa respectively (Vuono et al., 2015). Its use, over Pearson correlation, was justified since all variable combinations, transformed as in 2.4.1, were not bivariate normal (roystonTest, "MVN" v. 4.0.2, Korkmaz et al. (2014)). Bonferroni corrections were applied to all calculated correlations.

2.4.3. Virus – community function interactions

Virus-community function interactions were assessed by calculation of Spearman's rank correlation coefficients between mixed liquor and effluent virus abundance and mixed liquor and effluent COD and $\text{NH}_4^+ - \text{N}$ concentrations respectively, performed, justified and corrected as in 2.4.1. Structured equation modelling (SEM) was also conducted to infer hypothesised causal links between virus abundance and effluent COD and $\text{NH}_4^+ - \text{N}$ concentrations (sem, "lavaan" v. 0.5–23.1097, Rosseel (2012)). *A priori* models were constructed based on literature and theory, improved upon using statistical associations identified by regression analysis (2.4.1) and, to guarantee parsimony, simplified by removing non-significant indicators and pathways ($P < 0.1$). Model fits were assessed using the χ^2 test ($P > 0.05$), the root square mean error of approximation (RMSEA, < 0.06), the root mean square residual (RMR, < 0.08), the comparative fit index (CFI, > 0.95) and AIC respectively (Hu and Bentler, 1999; Grace et al., 2010), with R^2 values obtained for each dependent matrix (Grace et al., 2010).

Prior to SEM analysis all bivariate relationships were checked for nonlinearity (cor.test, "stats" v. 3.4.0, R Core Team, 2017) and, through evaluation of skewness and kurtosis values, multivariate normality (mardiaTest, "MVN" v. 4.0.2, Korkmaz et al., 2014). To aid in adherence to these assumptions variables were transformed (\log_{10} and natural log for biotic and abiotic respectively) and subsequently standardised to mean 0 variance 1, whilst SEM models, to counter minor normality departures, were estimated using a robust maximum likelihood approach (estimator = "MLM", "lavaan" v. 0.5–23.1097, Rosseel, 2012) and Satorra-Bentler scaled fit statistics (Curran et al., 1996).

3. Results

3.1. Bioreactor performance and abiotic conditions

Influent flow rates were highly variable over the two-year study period (Table S1 and Fig. S1), leading to fluctuating hydraulic and solids retention times (10.6 ± 3.1 h and 11.1 ± 3.1 days respectively, Fig. S1) and reactor biomass concentrations (2.7 ± 0.5 g L⁻¹ MLSS and 2.01 ± 0.4 g L⁻¹ MLVSS, Fig. S2). Reactor temperature varied on a seasonal basis, with a summer maximum and winter minimum of 17.9 °C (August 2012) and 6.3 °C (January 2013) recorded respectively (Fig. S1). DO concentrations were moderately variable (2.5 ± 0.8 mg L⁻¹) and pH was maintained within a narrow range (6.6 ± 0.2) through flow dependent dosing of sodium hydroxide (Fig. S1).

The discharge consent, a legislated maximum final effluent concentration, of COD_s (125 mg L⁻¹) was achieved 96.1% of the time, thus mean removal efficiency ($82.1 \pm 18.1\%$) and effluent concentrations (32 ± 28.5 mg L⁻¹) were relatively stable (Fig. S2 B). Nitrification was less efficient and more unstable (mean $\text{NH}_4^+ - \text{N}$ removal of $75.8 \pm 27.6\%$); accordingly effluent $\text{NH}_4^+ - \text{N}$ and nitrite concentrations were highly variable (0–76.4 and 0–5.3 mg L⁻¹) and the discharge consent (5 mg L⁻¹ ammonia) was adhered to only 57.8% of the time (Fig. S2). Certainly nitrite accumulation, thus incomplete nitrification, was periodically evident, particularly from day 500 onwards (Fig. S2 E). Of the trace metals monitored in the influent calcium, potassium and magnesium were present at the highest concentrations (mean values of 47.4 ± 8.7 , 13.1 ± 3.1 and 11 ± 2.8 mg L⁻¹ respectively), whilst cadmium and arsenic were present at the lowest concentrations (mean values of 3.9 ± 11.3 and 9.3 ± 7.9 µg L⁻¹ respectively, Fig. S3). All monitored trace metals, however, were routinely present in the influent (Table S1 and Fig. S3).

3.2. Temporal abundance dynamics of viruses and bacteria

Mixed liquor virus abundance varied by roughly an order of magnitude (3.18×10^8 – 3.41×10^9 viruses mL⁻¹) throughout the two years, with temporal fluctuations evident (Fig. 1 A). Virus abundance was greater than bacterial abundance; averaging $1.31 \pm 0.57 \times 10^9$ viruses mL⁻¹ compared with $4.30 \pm 3.48 \times 10^8$ bacteria mL⁻¹ and $1.36 \pm 1.13 \times 10^7$ AOB mL⁻¹. The abundance of total bacteria and AOB also varied across orders of magnitude (4.29×10^7 – 2.20×10^9 bacteria mL⁻¹ and 1.69×10^6 – 6.04×10^7 AOB mL⁻¹), with fluctuations generally, but not exclusively, occurring concurrently (Fig. 1 B and C). The mean mixed liquor virus to total bacteria ratio (VBR) was 6.15, although this varied greatly (0.52–76.86).

The abundance of all three communities within the influent was orders of magnitude lower (Fig. 1), with mean abundances of $1.52 \pm 0.83 \times 10^8$ viruses mL⁻¹, $2.35 \pm 4.24 \times 10^7$ bacteria mL⁻¹ and $1.14 \pm 2.98 \times 10^5$ AOB mL⁻¹ recorded respectively. Synchronous, order of magnitude abundance fluctuations, analogous to those in the mixed liquor, were again evident (1.79×10^6 – 4.38×10^8 viruses mL⁻¹, 2.81×10^5 – 23.62×10^8 bacteria mL⁻¹ and 2.1×10^3 – 2.8×10^6 AOB mL⁻¹, Fig. 1). Effluent virus abundance ($6.1 \pm 2.39 \times 10^8$ viruses mL⁻¹) was also highly variable (1.1×10^8 – 1.27×10^9 viruses mL⁻¹, Fig. 1 A) and often mimicked abundance fluctuations in the mixed liquor. Indeed mixed liquor and effluent virus abundance were positively correlated ($P < 0.001$, Fig. 1), as viruses in the mixed liquor were lost in effluent wastewater.

3.3. Virus interactions with biotic and abiotic conditions

Fitted values from the mixed liquor virus GLS regression model

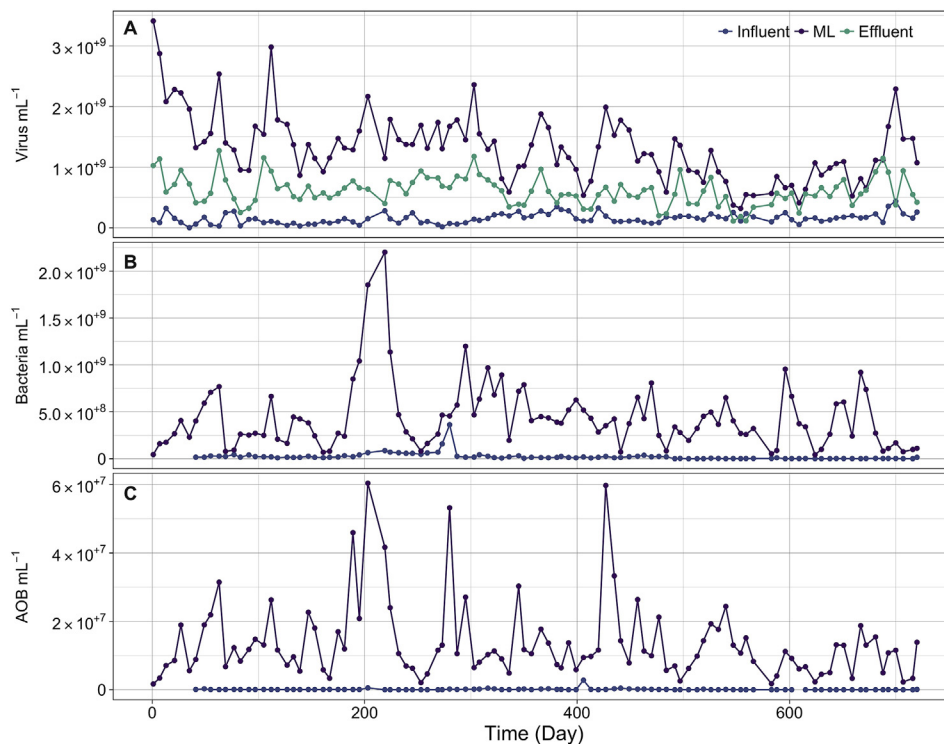


Fig. 1. The abundance of viruses (A), total bacteria (B) and AOB (C) in the influent, mixed liquor and effluent (viruses only) of the full scale WWTP. ML = mixed liquor.

agreed well with observed abundances (Fig. 2 A and Fig. S4 A), with explanatory variables explaining 83% of abundance variations. The model identified strong positive associations with influent virus abundance ($P < 0.05$), influent NH_4^+ -N ($P < 0.01$), phosphate

($P < 0.01$) and sulphur ($P < 0.001$) concentrations, mixed liquor AOB abundance ($P < 0.05$) and mixed liquor pH ($P < 0.05$). In contrast influent magnesium concentrations ($P < 0.001$) and mixed liquor nitrate ($P < 0.001$), nitrite ($P < 0.01$) and sulphate ($P < 0.001$)

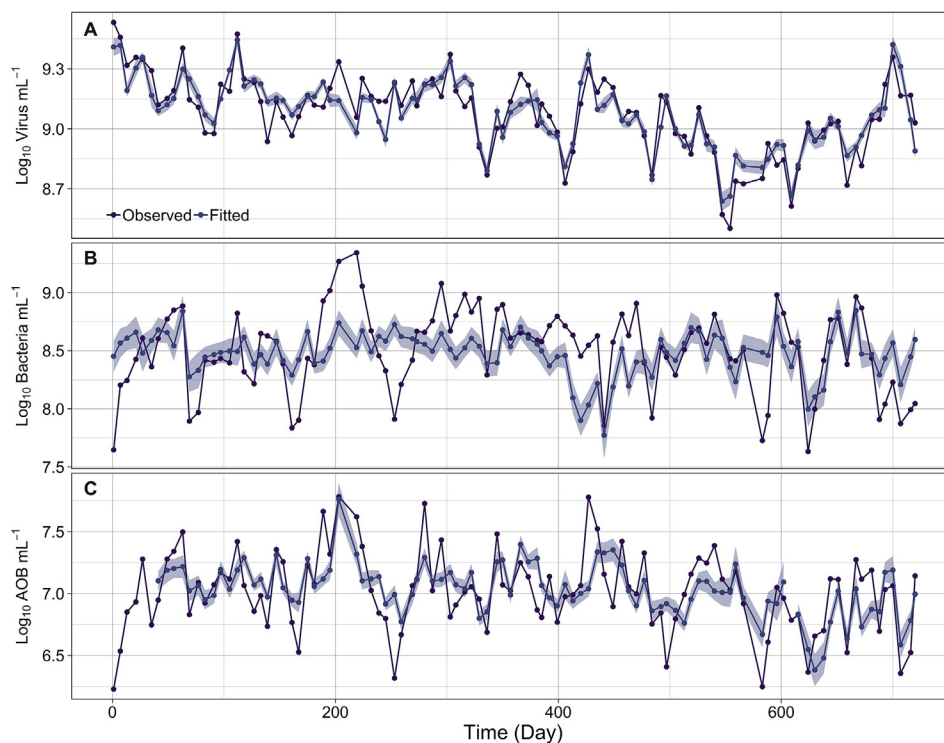


Fig. 2. \log_{10} mixed liquor virus (A), bacteria (B) and AOB (C) abundance from the full scale WWTP (observed) and respective multivariate GLS regression models (fitted). Shaded area represents 95% confidence intervals for the regression models. $n = 102$ (A and B) and 95 (C) respectively, the latter due to missingness in influent AOB abundance.

concentrations were highly negatively associated (Table S2) with total virus numbers.

However, the fits of the mixed liquor total bacteria and AOB GLS regression models, when compared to that produced for mixed liquor viruses, were poorer (Fig. 2, Fig S5 A and S6 A). The models explained only 53% and 47% of the variation in the bacterial and AOB abundances respectively. However mixed liquor virus abundances had strong positive associations with both bacteria and AOB concentrations ($P < 0.01$ and $P < 0.001$ respectively Tables S3 and S4). The AOB were even positively correlated with viruses in the effluent ($P < 0.01$). Other explanatory variables significantly contributing to each model are summarised in the supplementary information.

3.4. Virus interactions with bacteria community structure

The CCA ordination accounted for the majority of observed variance in OTU-environment associations (77.6%, Table S5 and Fig. 3). Mixed liquor virus abundance ($P < 0.001$) was the third most important variable after water temperature, $P < 0.001$, and influent potassium, $P < 0.001$ (Fig. 3 and Table S5).

Mixed liquor virus abundance was found to have a significant ($LS > 0.3$, $P \leq 0.05$ and $Q \leq 0.01$) and complex association with 6 of the 10 most abundant OTUs and a further 11 of the 50 most abundant taxa (Fig. 4). A group of 5 taxa had a significant association with the virus concentration at the time of sampling (both negative, OTUs 8 and 24, and positive OTUs 5, 9 and 26). This group of OTUs were highly interconnected, OTUs 5, 9 and 26 being positively associated and OTU pairs 8 and 5, 8 and 9, 8 and 26 and 24 and 9 being negatively associated (Fig. 4), all associations were non-lagged. All 5 were persistent in the mixed liquor (PAR ≤ 5 , Fig. 4) and included the “workhorse” genus *Zoogloea* (OTU 5).

However, other taxa had associations with a time delay, varying from 1 week (OTU 11 positive and OTUs 33, 37 and 44 negative), 2 weeks (positive, OTUs 2 and 50 and negative, OTUs 5, 34 and 46)

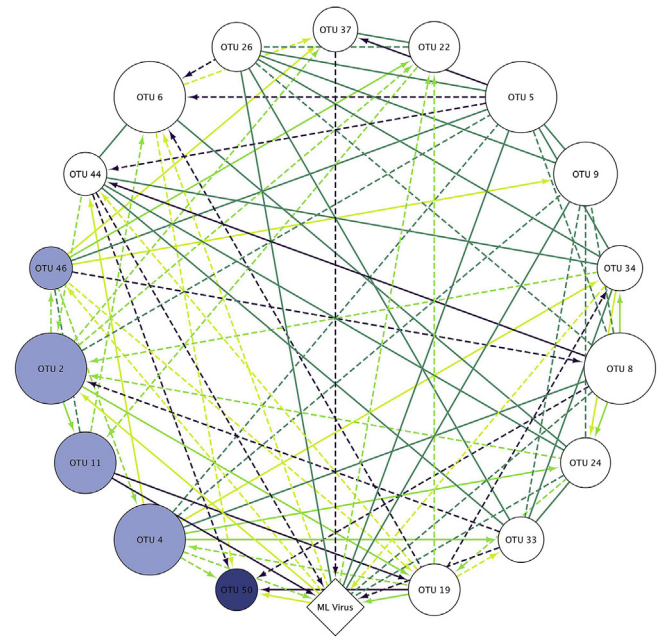


Fig. 4. Network visualisation of local similarity analysis associations between all OTUs from the top 50 most abundant that were associated with mixed liquor virus abundance ($LS \geq 0.3$, $P \leq 0.05$ and $Q \leq 0.01$) at the full scale WWTP. Circular nodes indicate bacterial taxa, coloured and sized based on the OTU's PAR (≤ 5 = white, $> 5 < 10$ = light blue and ≥ 10 = dark blue) and mean relative abundance respectively. For reference OTU 2 and OTU 50 had mean relative abundances of 4.68%, the highest, and 0.36%, the lowest. The diamond shaped node indicates mixed liquor virus abundance. Solid and dashed lines indicate positive and negative associations respectively, whilst a lines colour indicates the correlations delay (teal = 0, purple = 1, yellow = 2 and green = 3). Arrows point to the delayed, or trailing, node in delayed associations. ML = mixed liquor.

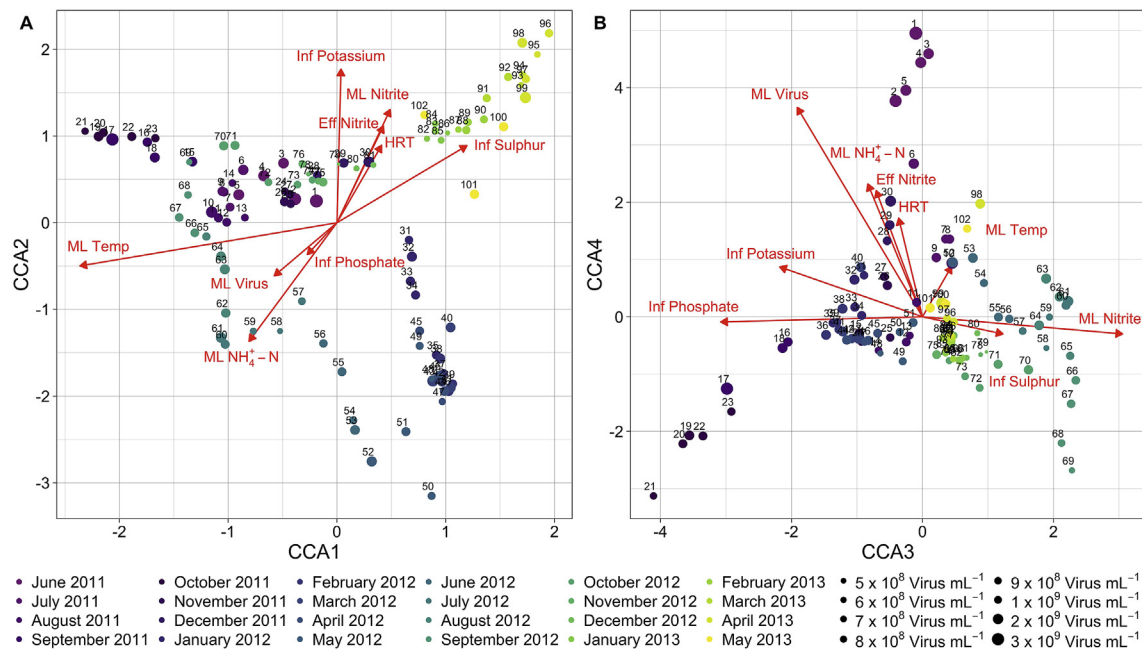


Fig. 3. Canonical correspondence analysis ordination of temporal mixed liquor bacterial community dynamics from the full scale WWTP and associated virus and abiotic variables, axes 1 and 2 (A) and 3 and 4 (B) respectively. Points and associated numbers represent individual samples, coloured and sized based on the month and year of collection and the samples virus concentration mL⁻¹ respectively. Proximity of samples indicates similarities in community composition. Arrows indicate the direction of increase in each explanatory variable, their length indicates the strength of correlation with each axis and the angle between arrows indicates the approximate degree to which explanatory variables are correlated. Inf = influent, ML = mixed liquor and Eff = effluent.

and even 3 weeks (positive, OTU 19 and negative, OTUs 4 and 22). These OTUs were found intermittently ($PAR > 5 < 10$) or transiently ($PAR \geq 10$) and included well known filamentous organisms, *Microthrix* (OTUs 2, 11 and 37), *Thiothrix* (OTU 50) and the phosphorous removing genus *Tetrasphaera* (OTU 4). Mixed liquor virus abundance was not statistically associated with either Hill's diversity indices (Fig. 5 A).

3.5. Virus interactions with community function

Mixed liquor and effluent virus abundance were highly positively correlated with both mixed liquor COD ($P < 0.001$ and $P < 0.01$) and $NH_4^+ - N$ ($P < 0.001$ and $P < 0.05$) concentrations, and weakly positively correlated with effluent $NH_4^+ - N$ concentrations ($P < 0.1$ and $P < 0.1$, Fig. 5 B).

The SEM also identified significant positive associations, although non-directional, between mixed liquor virus abundance and both mixed liquor $NH_4^+ - N$ ($P < 0.05$, Fig. 6 A) and COD ($P < 0.01$, Fig. 6 B) concentrations, which were directly influencing, and being influenced by, corresponding effluent ($P < 0.001$ and $P < 0.01$, Fig. 6) and influent concentrations ($P < 0.001$ and $P < 0.05$, Fig. 6) respectively. Effluent virus abundance was also none directionally positively associated with effluent $NH_4^+ - N$ ($P < 0.05$, Fig. 6 A) and COD ($P < 0.01$, Fig. 6 B) concentrations, whilst being directly positively influenced by mixed liquor virus abundance ($P < 0.001$ and $P < 0.001$, Fig. 6). Mixed liquor virus abundance ($P < 0.001$, Fig. 6 A), as well as mixed liquor AOB abundance ($P < 0.1$, Fig. 6 A), was also directly positively influenced by influent concentrations of $NH_4^+ - N$. No significant associations were found between mixed liquor AOB/bacteria abundance and mixed liquor and effluent $NH_4^+ - N$ /COD concentrations and virus abundance (Fig. 6), the only exception being the negative association found between mixed liquor AOB and effluent virus abundance ($P < 0.01$, Fig. 6 A). Overall the SEM analysis was able to explain 28% and 12% of effluent $NH_4^+ - N$ and COD concentrations respectively.

4. Discussion

This work provides evidence that viruses play an important role in the ecology and function of activated sludge plants: the statistical associations imply they interact with the abundance and dynamics

of key bacterial groups to such an extent that community function, and therefore effluent quality is affected. It seems plausible that viruses could cause outright failure either directly, affecting key functional groups, or indirectly by influencing filamentous bacteria and thus foaming and bulking.

Such a prominent role for bacteriophage is consistent with the high virus to bacteria ratio we observed, evidence from other engineered systems (Hantula et al., 1991; Lee et al., 2007; Barr et al., 2010; Shapiro et al., 2010) and knowledge of their role in the ecology of marine and freshwater environments (Suttle, 2007; Rohwer and Thurber, 2009; Breitbart, 2012), where viruses are thought to kill ~50% of bacteria every day (Noble and Fuhrman, 2000; Ram et al., 2013) and therefore influence nutrient cycling and community function.

Inferring the role of viruses in the environment is however problematic; we are only able to quantify total virus numbers and the abundance or proportional abundance of relatively broad bacterial groups. Moreover, the virus community will comprise numerous species/strains (Breitbart et al., 2007), each with varying host ranges (Khan et al., 2002a, 2002b) and most will likely be in unsynchronised predator prey cycles with their hosts. In addition, the relationship between phage and host may be non-linear and lagged.

Thus in the absence of more detailed measurements of viruses and their hosts we utilised multiple statistical approaches to identify significant relationships between total virus numbers and various measures of biomass, community structure and function. This is a crude and probably insensitive approach. However, if we can find significant relationships with relatively weak tests, this does imply that viruses play a very strong role in this system. Likewise, though the underlying basis of the observed correlations is uncertain, and could be indirect, they could provide the impetus for the more detailed and challenging work required to obtain a mechanistic understanding of how bacteriophage affect engineered biological systems.

All four lines of enquiry we used suggested a role for viruses (regression and structural equation modelling to determine an impact on biomass and function, effluent quality, and CCA and local similarity analysis to evaluate the effect of viruses on community structure). A detailed consideration of every statistically significant relationship would be tedious and speculative. Nevertheless some

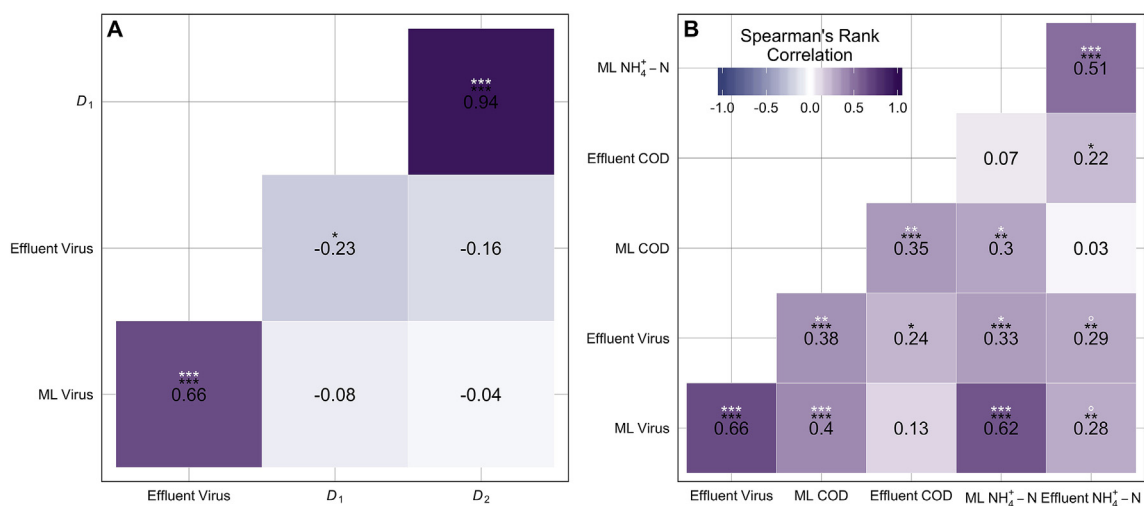


Fig. 5. Spearman's rank correlation coefficients describing virus – community structure (A) and virus – community function (B) interactions. Black and white α values are standard and Bonferroni corrected respectively. Royston's Multivariate Normality Test: $P > 0.05$ for all correlations and $n = 102$ (A and B). ML = mixed liquor, D_1 = exponential of Shannon diversity and D_2 = inverse of Simpson diversity. ° $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

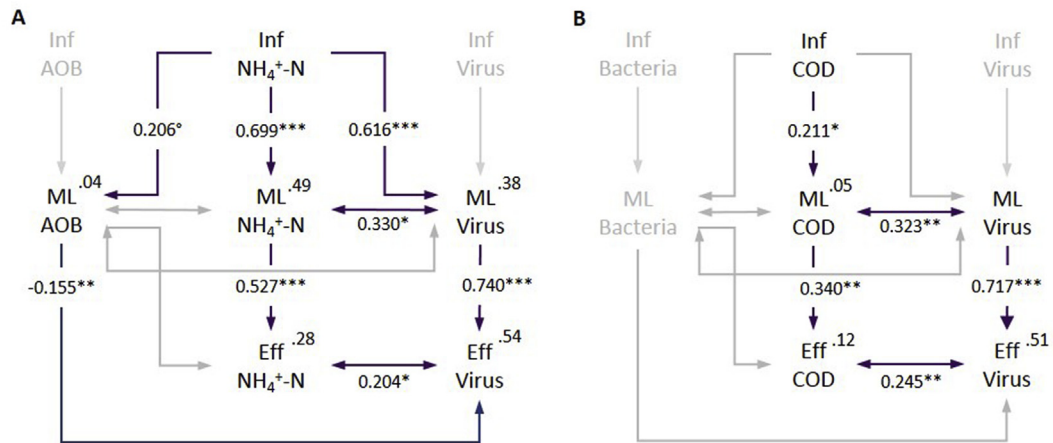


Fig. 6. Structured equation models describing viral interaction with $\text{NH}_4^+\text{-N}$ (A) and COD (B) concentrations throughout the treatment stream at the full scale WWTP. Standardised path coefficients and their significance are given along path arrows, which are coloured based on sign and significance (purple = positive, blue = negative, grey = non-significant (removed from final model)). Double-headed arrows indicate associations lacking a clear direction of causality, whilst the explained variance (R^2) of each endogenous variable is proximal to its label. $n = 102$, $\chi^2 = 7.79$ (A, $P = 0.352$) and 3.64 (B, $P = 0.603$), RMSEA = 0.033 (A) and 0.000 (B), RMR = 0.044 (A) and 0.052 (B) and CFI = 0.997 (A) and 1.000 (B). ° $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Inf = influent, ML = mixed liquor and Eff = effluent.

commentary is warranted.

Viruses emerged as having a significant, and positive correlation with the variation in bacterial and AOB biomass over time. More biomass (prey) meant more viruses (predator), which is plausible. In individual predator-prey cycles the predator lags the prey but will still be positively correlated. It would appear that when many such cycles are aggregated this correlation is still detectable. The overall fit of the regression models were modest and the mechanistic basis for many of the other statistically significant variables is not immediately apparent. However, some have an obvious physical basis that lends credence to the exercise. For example pH and the divalent cation magnesium were associated with virus numbers: both regulate surface charge and electrostatic interactions and would cause viruses to adsorb/desorb from the biomass and so decrease/increase their abundance (Michen and Graule, 2010; Nap et al., 2014). Likewise viruses in the mixed liquor were influenced by viruses in the influent, despite the latter constituting only a small fraction of the former. The difference between the number of viruses in the influent and the mixed liquor can only be accounted for by the lysis of biomass in the mixed liquor. Yet though the time series of mixed liquor virus abundance could be modelled quite well there was no statistical relationship between the number of viruses and the amount of bacterial biomass. This reinforces the limitations of a purely statistical approach.

In principle, if viruses affect mixed liquor AOB and bacterial biomass, this in turn should affect the effluent COD and ammonia concentration and so the system's function. The structural equation modelling was an attempt to show this chain of causality. It partially succeeded. Though it is clear that viruses affected the effluent concentrations directly and indirectly it was not possible to demonstrate that they did this through a linear effect on bacterial concentrations. However, predation remains the most plausible mechanism by which viruses could affect effluent quality. Indeed key functional groups are known to have signatures of viral infection (prophages and CRISPR sequences) in their genomes (Kunin et al., 2008; Choi et al., 2010; Motlagh et al., 2015).

Viruses appear to have an important, possibly very important, role in community structure being the third most important factor in the canonical correspondence analysis (CCA) after temperature and potassium. The importance of temperature is well established

and to some extent obvious (Wells et al., 2009, 2011), as is the significance of potassium (which constitute about 1% of cell biomass, e.g. Brdjanovic et al. (1996) and Barat et al. (2005)). It follows that the prominent role of viruses is credible though largely ignored in both research and practice and poorly understood in detail. Unfortunately though a CCA can tell us what is important: it cannot tell us why.

Some preliminary insights can be garnered from the local similarity analysis of individual taxonomic groups. Significant associations were found between viruses and some of the most abundant taxa, which is consistent with the Killing the Winner hypothesis (KtW, Thingstad (2000); Winter et al. (2010)). However, given viruses of the most abundant taxa should constitute the bulk of the virus count such relationships are perhaps expected. Thus it is in some ways more remarkable that total viruses were seen to correlate with *Microthrix*, *Thiothrix* and *Tetrasphaera* genera, problematic blooming filamentous bacteria (Blackall et al., 1996; Martins et al., 2004; Guo and Zhang, 2012) that cause foaming and bulking and whose abundance is a matter of great interest and importance to practitioners. A more prosaic explanation could be that some deterministic factor, for example pH, could be affecting both viruses and these problematic bacteria, leading to spurious correlations. However, our observations are consistent with findings elsewhere (Kotay et al., 2011; Petrovski et al., 2012; Khairnar et al., 2014; Liu et al., 2015b). The time lag in some of the relationships we observe probably reflect the asynchrony in the underlying dynamics.

5. Conclusion

We conclude that this paper provides plausible evidence that bacteriophage are key players in the ecology of activated sludge, affecting not only the dynamics of the biomass and the community structure but the function of the system: carbon and ammonia removal. We posit that bacteriophages cause failure: either directly through effluent concentrations or indirectly through their interactions with filamentous organisms. There seems no reason why this finding will not extend to other functions (especially P removal) and all other biological treatment systems. We caution that all findings are based on statistical associations and their basis is not yet certain. Clearly, deeper mechanistic research is needed to

determine the detailed, species specific, mechanisms. This endeavour will be strengthened by the ongoing revolution in viral metagenomics and single cell analysis (Dang and Sullivan, 2014; Brum and Sullivan, 2015). However, practitioners need not wait for this detailed work to be undertaken. The quantification of bacteriophage is cheap and simple using flow cytometry and total virus counts can, as we have shown, be empirically related to plant performance. We would suggest that operators initiate the use of flow cytometry to count viruses (or VLP) in treatment plants now.

Acknowledgements

This work was carried out as part of a Frontiers in Engineering research project (NUFEB, <http://research.ncl.ac.uk/nufeb/>) awarded and funded by the Engineering and Physical Sciences Research Council (EP/H012133/1). We thank Northumbrian Water who permitted collection of wastewater samples from the full scale plant. Data supporting this publication is openly available under an 'Open Data Commons Open Database License' and is available at <https://doi.org/10.17634/121736-6>.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.watres.2019.01.003>.

References

- APHA, 1998. Standard Methods for the Examination of Water and Wastewater. APHA.
- Ayarza, J.M., Erijman, L., 2011. Balance of neutral and deterministic components in the dynamics of activated sludge floc assembly. *Microb. Ecol.* 61 (3), 486–495.
- Ayarza, J.M., Guerrero, L.D., Erijman, L., 2010. Nonrandom assembly of bacterial populations in activated sludge flocs. *Microb. Ecol.* 59 (3), 436–444.
- Barat, R., Montoya, T., Seco, A., Ferrer, J., 2005. The role of potassium, magnesium and calcium in the enhanced biological phosphorus removal treatment plants. *Environ. Technol.* 26 (9), 983–992.
- Barr, J.J., Slater, F.R., Fukushima, T., Bond, P.L., 2010. Evidence for bacteriophage activity causing community and performance changes in a phosphorus-removal activated sludge. *FEMS Microbiol. Ecol.* 74 (3), 631–642.
- Barton, K., 2016. MuMIn: Multi-Model Inference.
- Bergh, O., Borsheim, K.Y., Bratbak, G., Haldal, M., 1989. High abundance of viruses found in aquatic environments. *Nature* 340 (6233), 467–468.
- Bettarel, Y., Sime-Ngando, T., Amblard, C., Dolan, J., 2004. Viral activity in two contrasting lake ecosystems. *Appl. Environ. Microbiol.* 70 (5), 2941–2951.
- Billar, S.J., McDaniel, L.D., Breitbart, M., Rogers, E., Paul, J.H., Chisholm, S.W., 2017. Membrane vesicles in sea water: heterogeneous DNA content and implications for viral abundance estimates. *ISME J.* 11 (2), 394–404.
- Blackall, L.L., Seviour, E.M., Bradford, D., Stratton, H.M., Cunningham, M.A., Hugenholtz, P., Seviour, R.J., 1996. Towards understanding the taxonomy of some of the filamentous bacteria causing bulking and foaming in activated sludge plants. *Water Sci. Technol.* 34 (5–6), 137–144.
- Bratbak, G., Haldal, M., Thingstad, T.F., Tuomi, P., 1996. Dynamics of virus abundance in coastal seawater. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol.* 19 (4), 263–269.
- Brđjanovic, D., Hooijmans, C.M., van Loosdrecht, M.C.M., Alaerts, G.J., Heijnen, J.J., 1996. The dynamic effects of potassium limitation on biological phosphorus removal. *Water Res.* 30 (10), 2323–2328.
- Breitbart, M., 2012. Marine Viruses: Truth or Dare. *Annual Review of Marine Science*, vol. 4. C. A. G. S. J. Carlson, pp. 425–448, 4.
- Breitbart, M.Y.A., Thompson, L.R., Suttle, C.A., Sullivan, M.B., 2007. Exploring the vast diversity of marine viruses. *Oceanography* 20 (2), 135–139.
- Brown, M.R., Camézuli, S., Davenport, R.J., Petelenz-Kurdiel, E., Øvreås, L., Curtis, T.P., 2015. Flow cytometric quantification of viruses in activated sludge. *Water Res.* 68 (0), 414–422.
- Brum, J.R., Sullivan, M.B., 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* 13 (3), 147–159.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6 (8), 1621–1624.
- Choi, J.D., Kotay, S.M., Goel, R., 2010. Various physico-chemical stress factors cause prophage induction in *Nitrosospora multiformis* 25196—an ammonia oxidizing bacteria. *Water Res.* 44 (15), 4550–4558.
- Curran, P.J., West, S.G., Finch, J.F., 1996. The robustness of test statistics to non-normality and specification error in confirmatory factor Analysis. *Psychol. Methods* 1 (1), 16–29.
- Dang, V.T., Sullivan, M.B., 2014. Emerging methods to study bacteriophage infection at the single-cell level. *Front. Microbiol.* 5, 724.
- Forterre, P., Soler, N., Krupovic, M., Marguet, E., Ackermann, H.W., 2013. Fake virus particles generated by fluorescence microscopy. *Trends Microbiol.* 21 (1), 1–5.
- Fox, J., Weisberg, S., 2011. An {R} Companion to Applied Regression, second ed. Sage, Thousand Oaks CA.
- Grace, J.B., Anderson, T.M., Olf, H., Scheiner, S.M., 2010. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* 80 (1), 67–87.
- Gross, J., Ligges, U., 2015. Nortest: Tests for Normality.
- Guo, F., Zhang, T., 2012. Profiling bulking and foaming bacteria in activated sludge by high throughput sequencing. *Water Res.* 46 (8), 2772–2782.
- Haaber, J., Middelboe, M., 2009. Viral lysis of *Phaeocystis pouchetii*: implications for algal population dynamics and heterotrophic C, N and P cycling. *ISME J.* 3 (4), 430–441.
- Hantula, J., Kurki, A., Vuoriranta, P., Bamford, D.H., 1991. Ecology of bacteriophages infecting activated sludge bacteria. *Appl. Environ. Microbiol.* 57 (8), 2147–2151.
- Hara, S., Terauchi, K., Koike, I., 1991. Abundance of viruses in marine waters - assessment by epifluorescence and transmission electron-microscopy. *Appl. Environ. Microbiol.* 57 (9), 2731–2734.
- Hennes, K.P., Simon, M., 1995. Significance of bacteriophages for controlling bacterioplankton growth in a mesotrophic lake. *Appl. Environ. Microbiol.* 61 (1), 333–340.
- Hofer, J.S., Sommaruga, R., 2001. Seasonal dynamics of viruses in an alpine lake: importance of filamentous forms. *Aquat. Microb. Ecol.* 26 (1), 1–11.
- Hu, L. t., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.: Multidiscip. J.* 6 (1), 1–55.
- Huber, J.A., Mark Welch, D., Morrison, H.G., Huse, S.M., Neal, P.R., Butterfield, D.A., Sogin, M.L., 2007. Microbial population structures in the deep marine biosphere. *Science* 318 (5847), 97–100.
- Jacquet, S., Domaizon, I., Personnic, S., Ram, A.S.P., Hedat, M., Duhamel, S., Sime-Ngando, T., 2005. Estimates of protozoan- and viral-mediated mortality of bacterioplankton in Lake Bourget (France). *Freshw. Biol.* 50 (4), 627–645.
- Jiang, S.C., Paul, J.H., 1994. Seasonal and diel abundance of viruses and occurrence of lysogeny/bacteriocinogeny in the marine-environment. *Mar. Ecol. Prog. Ser.* 104 (1–2), 163–172.
- Khairnar, K., Pal, P., Chandekar, R.H., Pauniar, W.N., 2014. Isolation and characterization of bacteriophages infecting nocardioforms in wastewater treatment plant. *Biotechnol. Res. Int.* 2014, 151952.
- Khan, M.A., Satoh, H., Katayama, H., Kurisu, F., Mino, T., 2002a. Bacteriophages isolated from activated sludge processes and their polyvalency. *Water Res.* 36 (13), 3364–3370.
- Khan, M.A., Satoh, H., Mino, T., Katayama, H., Kurisu, F., Matsuo, T., 2002b. Bacteriophage-host interaction in the enhanced biological phosphate removing activated sludge system. *Water Sci. Technol.* 46 (1–2), 39–43.
- Korkmaz, S., Goksuluk, D., Zararsiz, G., 2014. MVN: an R package for assessing multivariate normality. *R J.* 6 (2), 151–162.
- Kotay, S.M., Datta, T., Choi, J., Goel, R., 2011. Biocontrol of biomass bulking caused by *Haliscomenobacter hydrossis* using a newly isolated lytic bacteriophage. *Water Res.* 45 (2), 694–704.
- Kunin, V., He, S., Warnecke, F., Peterson, S.B., Garcia Martin, H., Haynes, M., Ivanova, N., Blackall, L.L., Breitbart, M., Rohwer, F., McMahon, K.D., Hugenholtz, P., 2008. A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res.* 18 (2), 293–297.
- Lee, S.H., Otawa, K., Onuki, M., Satoh, H., Mino, T., 2007. Population dynamics of phage-host system of *Microcylindrus phosphovorus* indigenous in activated sludge. *J. Microbiol. Biotechnol.* 17 (10), 1704–1707.
- Li, W.K.W., Dickie, P.M., 2001. Monitoring phytoplankton, bacterioplankton, and viroplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry* 44 (3), 236–246.
- Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., Chisholm, S.W., 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U. S. A.* 101 (30), 11013–11018.
- Liu, H., Yuan, X., Xu, J., Harrison, P.J., He, L., Yin, K., 2015a. Effects of viruses on bacterial functions under contrasting nutritional conditions for four species of bacteria isolated from Hong Kong waters. *Sci. Rep.* 5, 14217.
- Liu, M., Gill, J.J., Young, R., Summer, E.J., 2015b. Bacteriophages of wastewater foaming-associated filamentous *Gordonia* reduce host levels in raw activated sludge. *Sci. Rep.* 5, 13754.
- Marie, D., Brussaard, C.P.D., Thyrhaug, R., Bratbak, G., Vaulot, D., 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl. Environ. Microbiol.* 65 (1), 45–52.
- Martin, T.D., Creed, J.T., Brockhoff, C.A., 1994. U.S. Environmental Protection Agency Report Revision 2. Sample Preparation Procedure for Spectrochemical Determination of Total Recoverable Elements, vol. 8, p. 12.
- Martins, A.M.P., Pagilla, K., Heijnen, J.J., van Loosdrecht, M.C.M., 2004. Filamentous bulking sludge—a critical review. *Water Res.* 38 (4), 793–817.
- McTavish, H., Fuchs, J.A., Hooper, A.B., 1993. Sequence of the gene coding for ammonia monooxygenase in *Nitrosomonas europaea*. *J. Bacteriol.* 175 (8), 2436–2444.
- Michen, B., Graule, T., 2010. Isoelectric points of viruses. *J. Appl. Microbiol.* 109 (2),

- 388–397.
- Middelboe, M., Jorgensen, N.O.G., 2006. Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. *J. Mar. Biol. Assoc. U. K.* 86 (3), 605–612.
- Motlagh, A.M., Bhattacharjee, A.S., Goel, R., 2015. Microbiological study of bacteriophage induction in the presence of chemical stress factors in enhanced biological phosphorus removal (EBPR). *Water Res.* 81, 1–14.
- Muyzer, G., de Waal, E.C., Uitterlinden, A.G., 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59 (3), 695–700.
- Nap, Rikkert J., Božić, Anže L., Szleifer, I., Podgornik, R., 2014. The role of solution conditions in the bacteriophage PP7 capsid charge regulation. *Biophys. J.* 107 (8), 1970–1979.
- Noble, R.T., Fuhrman, J.A., 2000. Rapid virus production and removal as measured with fluorescently labeled viruses as tracers. *Appl. Environ. Microbiol.* 66 (9), 3790–3797.
- Norton, J., Alzerreca, J., Suwa, Y., Klotz, M., 2002. Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch. Microbiol.* 177 (2), 139–149.
- Ofiteru, I.D., Lunn, M., Curtis, T.P., Wells, G.F., Criddle, C.S., Francis, C.A., Sloan, W.T., 2010. Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci. U. S. A.* 107 (35), 15345–15350.
- Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoezs, E., Wagner, H., 2018. *Vegan: Community Ecology Package*.
- Otawa, K., Lee, S.H., Yamazoe, A., Onuki, M., Satoh, H., Mino, T., 2007. Abundance, diversity, and dynamics of viruses on microorganisms in activated sludge processes. *Microb. Ecol.* 53 (1), 143–152.
- Petrovski, S., Tillett, D., Seviour, R.J., 2012. Isolation and complete genome sequence of a bacteriophage lysing *Tetrasphaera jenkinsii*, a filamentous bacteria responsible for bulking in activated sludge. *Virus Gene.* 45 (2), 380–388.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkur, D., Team, R.C., 2017. *nlme: Linear and Nonlinear Mixed Effects Models*.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41 (D1), D590–D596.
- R Core Team, 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ram, A.S.P., Palesse, S., Colombet, J., Sabart, M., Perriere, F., Sime-Ngando, T., 2013. Variable viral and grazer control of prokaryotic growth efficiency in temperate freshwater lakes (French massif central). *Microb. Ecol.* 66 (4), 906–916.
- Ramette, A., 2007. Multivariate analyses in microbial ecology. *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Ecol.* 62 (2), 142–160.
- Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pasic, L., Thingstad, T.F., Rohwer, F., Mira, A., 2009. OPINION Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* 7 (11), 828–836.
- Rohwer, F., Thurber, R.V., 2009. Viruses manipulate the marine environment. *Nature* 459 (7244), 207–212.
- Rosseel, Y., 2012. *Lavaan: an R package for structural equation modeling*. *J. Stat. Softw.* 48 (2), 36, 2012.
- Rotthauwe, J.H., Witzel, K.P., Liesack, W., 1997. The ammonia monooxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl. Environ. Microbiol.* 63 (12), 4704–4712.
- Ruan, Q., Dutta, D., Schwalbach, M.S., Steele, J.A., Fuhrman, J.A., Sun, F., 2006. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* 22 (20), 2532–2538.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504.
- Shapiro, O.H., Kushmaro, A., Brenner, A., 2010. Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *ISME J.* 4 (3), 327–336.
- Shelford, E.J., Middelboe, M., Moller, E.F., Suttle, C.A., 2012. Virus-driven nitrogen cycling enhances phytoplankton growth. *Aquat. Microb. Ecol.* 66 (1), 41–46.
- Stephen, J.R., Chang, Y.J., Macnaughton, S.J., Kowalchuk, G.A., Leung, K.T., Flemming, C.A., White, D.C., 1999. Effect of toxic metals on indigenous soil beta-subgroup proteobacterium ammonia oxidizer community structure and protection against toxicity by inoculated metal-resistant bacteria. *Appl. Environ. Microbiol.* 65 (1), 95–101.
- Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. Unit. States Am.* 100 (16), 9440–9445.
- Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., Chisholm, S.W., 2006. Prevalence and evolution of Core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* 4 (8), e234.
- Suttle, C.A., 2007. Marine viruses - major players in the global ecosystem. *Nat. Rev. Microbiol.* 5 (10), 801–812.
- Tchobanoglous, G., Burton, F.L., Stensel, H.D., Metcalf, Eddy, I., 2003. *Wastewater Engineering: Treatment and Reuse*. McGraw-Hill Education.
- Thingstad, T.F., 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* 45 (6), 1320–1328.
- Valentin-Vargas, A., Toro-Labrador, G., Massol-Deyá, A.A., 2012. Bacterial community dynamics in full-scale Activated sludge bioreactors: operational and ecological factors driving community assembly and performance. *PLoS One* 7 (8) e42524.
- van der Gast, C.J., Ager, D., Lilley, A.K., 2008. Temporal scaling of bacterial taxa is influenced by both stochastic and deterministic ecological factors. *Environ. Microbiol.* 10 (6), 1411–1418.
- Venables, W.N., Ripley, B.D., Venables, W.N., 2002. *Modern Applied Statistics with S*. Springer, New York.
- Vuono, D.C., Benecke, J., Henkel, J., Navidi, W.C., Cath, T.Y., Munakata-Marr, J., Spear, J.R., Drewes, J.E., 2015. Disturbance and temporal partitioning of the activated sludge metacommunity. *ISME J.* 9 (2), 425–435.
- Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73 (16), 5261–5267.
- Weinbauer, M.G., Fuks, D., Puskaric, S., Peduzzi, P., 1995. Diel, seasonal, and depth-related variability of viruses and dissolved DNA in the northern Adriatic Sea. *Microb. Ecol.* 30 (1), 25–41.
- Wells, G.F., Park, H.D., Eggleston, B., Francis, C.A., Criddle, C.S., 2011. Fine-scale bacterial community dynamics and the taxa-time relationship within a full-scale activated sludge bioreactor. *Water Res.* 45 (17), 5476–5488.
- Wells, G.F., Park, H.D., Yeung, C.H., Eggleston, B., Francis, C.A., Criddle, C.S., 2009. Ammonia-oxidizing communities in a highly aerated full-scale activated sludge bioreactor: betaproteobacterial dynamics and low relative abundance of Crenarchaea. *Environ. Microbiol.* 11 (9), 2310–2328.
- Wigington, C.H., Sonderegger, D., Brussaard, C.P.D., Buchan, A., Finke, J.F., Fuhrman, J.A., Lennon, J.T., Middelboe, M., Suttle, C.A., Stock, C., Wilson, W.H., Wommack, K.E., Wilhelm, S.W., Weitz, J.S., 2016. Re-examination of the relationship between marine virus and microbial cell abundances. *Nat. Microbiol.* 1, 15024.
- Winter, C., Bouvier, T., Weinbauer, M.G., Thingstad, T.F., 2010. Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited. *Microbiol. Mol. Biol. Rev.* 74 (1), 42–57.
- Wommack, K.E., Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64 (1), 69–114.
- Wu, Q.L., Liu, W.T., 2009. Determination of Virus abundance, diversity and distribution in a municipal wastewater treatment plant. *Water Res.* 43 (4), 1101–1109.
- Xia, L.C., Ai, D., Cram, J., Fuhrman, J.A., Sun, F., 2013. Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. *Bioinformatics* 29 (2), 230–237.
- Xia, L.C., Steele, J.A., Cram, J.A., Cardon, Z.G., Simmons, S.L., Vallino, J.J., Fuhrman, J.A., Sun, F., 2011. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Syst. Biol.* 5 (2), S15.
- Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. *Meth. Ecol. Evol.* 1 (1), 3–14.